Australian Bureau of Statistics

2940.0.55.002 - Information Paper: Measuring Overcount and Undercount in the 2016 Population Census, Jul 2016

Latest ISSUE Released at 11:30 AM (CANBERRA TIME) 01/07/2016 First Issue

Summary

Introduction

INTRODUCTION

THE CENSUS OF POPULATION AND HOUSING

The Census is a valuable data source for estimating the size and geographic distribution of Australia's population, and for analysing the major demographic, social and economic characteristics by governments, businesses, community organisations and individuals. It is particularly useful for examining small geographic regions and small population groups. The Census also provides a base for post-censal population estimates and projections, which assist in planning and policy-making at the national, state and local level.

A Census is conducted in Australia every five years. The next Census will be held on the night of Tuesday 9 August 2016. On Census night, every person present in Australia, excluding foreign diplomats and their families, should be included on a Census form at the place where they stayed.

OVERCOUNT AND UNDERCOUNT

Whenever a Census is undertaken, questions about the completeness and accuracy of the Census count invariably arise. With such a large and complex exercise, it is inevitable that some people will be counted more than once and some may be missed.

Some of the reasons why people are counted more than once (i.e. overcounted) include:

- they were included on the Census form at the dwelling where they usually live, even though they stayed and were counted elsewhere on Census night; and
- they were overseas on Census night and so should not have been counted at all, but were included on the Census form at the dwelling where they usually live.

Some of the reasons why people may be missed (i.e. undercounted) include:

- they were travelling and were difficult to contact;
- they mistakenly thought they were counted elsewhere;
- there was insufficient space on the Census form in the dwelling where they were staying and additional forms were not obtained;
- the person completing the form thought that, for example, young babies, the elderly or visitors should not be included;

- they did not wish to be included due to reluctance to participate;
- the dwelling in which they were located was missed because it was difficult to find (e.g. in a remote or non-residential area); and
- the dwelling in which they were located was mistakenly classed as unoccupied.

After the Census collection period is finished, there are usually a small proportion of dwellings for which no Census form has been submitted. If Census fieldwork determines these dwellings are occupied, field information is used to impute a plausible number of persons present on Census night. As this imputation is approximate, it can also contribute to overcount or undercount.

POST ENUMERATION SURVEY

While every effort is made to eliminate these potential causes of error, some overcount and undercount will occur. The ABS obtains estimates of overcount and undercount using information collected in the Post Enumeration Survey (PES), which is conducted immediately following the Census.

If the PES finds that the total number of persons counted more than once in the Census is greater than the number of persons missed, the difference is called a net overcount. If the PES finds that the number of persons missed by the Census is greater than the number who were counted more than once, the difference is called a net undercount. Rates of net overcount or net undercount can vary significantly for different population groups depending on factors such as sex, age, ethnicity (including Indigenous origin) and geographic location.

Estimates of overcount and undercount are used to:

- derive an estimate of the resident population for 30 June of the Census year;
- provide users with an assessment of the completeness of Census counts, allowing them to take this into account when using Census information; and
- evaluate the effectiveness of Census collection procedures so that improvements can be made for future Censuses.

Accurate resident population estimates are required for a wide range of uses, including the allocation to states and territories of seats in the Federal House of Representatives; the distribution of Commonwealth payments to states and territories; planning and reporting on performance; and demographic, social and economic studies.

PURPOSE OF THIS PAPER

The purpose of this paper is to set out the methods to be used in the conduct of the 2016 Census Post Enumeration Survey.

Census Post Enumeration Survey

CENSUS POST ENUMERATION SURVEY

OVERVIEW

In Australia, estimates of overcount and undercount in the Census are based on the results of the Census Post Enumeration Survey (PES). The purpose of the PES is to determine how many people were counted more than once in the Census (and how many times these people were counted) and how many were missed.

The PES is a household survey conducted by specially trained interviewers starting about six weeks after Census night. This is a different collection methodology to the Census, where most forms are completed by any responsible adult on behalf of the household. A major advantage of interviewer-administered questionnaires is that people can be provided with assistance if they are uncertain about the meaning of any questions.

The scope of the PES includes all private dwellings in all areas, including Discrete Communities, but excluding special dwellings. For each dwelling selected in the PES, the any responsible adult (ARA) methodology is used as it is in Census. This generally involves a single person in the dwelling responding on behalf of all persons present or usually resident. In addition to obtaining basic demographic information, questions are asked about each person's usual residence, location on Census night, and any other addresses where they might have been counted in the Census.

Using this address information, the corresponding Census forms are examined to confirm how many times each respondent in the PES was counted in the Census. The results of this process are then combined and weighted to produce an estimate of net overcount or net undercount in the Census.

In the 2011 PES a sample of 36,000 private dwellings responded to the PES, about half of 1% of all private dwellings in Australia. The total number of people included in the PES was 94,000. The sampling fraction varied between states and territories, with larger sampling fractions in the states and territories with smaller populations. This design provides reliable estimates for each state and territory, as well as for Australia as a whole. Sampling fractions are also larger in Discrete Communities compared to the rest of the state they are in to ensure adequate reliability of Aboriginal and Torres Strait Islander estimates.

The 2016 PES sample size is larger than that for 2011, with a target of 43,000 responses from private dwellings. Additional information on the 2016 sample is provided in the Sample Design section of this paper.

Independence from the Census

INDEPENDENCE FROM THE CENSUS

One of the purposes of the PES is to provide an independent check on Census coverage. There are two aspects to this independence: operational independence and population independence.

Operational independence requires that Census operations do not influence PES operations and vice versa. Steps are taken to maintain the operational independence at every stage of the PES, including enumeration, processing and administration. These steps include:

- selecting the PES sample from an independent sample frame;
- using separate office staff in the PES and Census where possible;
- using a separate and secure information technology environment for processing the

PES information;

- ensuring the PES interviewers are not employed as Census field staff in the same area, and vice versa; and
- maintaining the confidentiality of the PES sample so that Census field and office staff are not aware which areas are included in the PES.

Population independence requires that people who were not counted in the Census are no more likely to be missed by the PES than people who were counted. In practice, some population subgroups may be less likely to respond to both the Census and the PES.

DIFFERENCES BETWEEN THE CENSUS AND THE PES

Householders are asked the PES questions using a personal interview by experienced, highly trained interviewers, whereas most Census forms are completed by a responsible adult on behalf of their household. The PES is also a much smaller scale operation (and hence easier to control) than the Census. These features enable the PES to deliver an accurate estimate of the percentage of people and dwellings missed or overcounted by the Census.

Just as the Census misses some people and dwellings that PES finds, so too the PES misses some people and dwellings that Census finds. In some cases this is because PES excludes certain kinds of persons and dwellings from scope for practical reasons, such as homeless persons and non-private dwellings (hotels, motels, hospitals). In other cases, PES misses people through non-contact or refusal of the occupants of a selected dwelling. The PES uses weight adjustment during estimation to account for the dwellings and people missed in the PES but counted in the Census.

LATE CENSUS FORMS

For some people who have not returned a Census form, contact from the ABS following selection in the PES acts as a reminder and possible motivator to return a completed Census form. These late returns, if not identified, would result in the PES sample having a higher proportion of Census response than in the overall population. To protect against this, all Census forms received after the start of PES enumeration are deemed 'late' and can't be used to improve accuracy of PES estimates.

CORRELATION BIAS

On the assumption that the Census and the PES are independent, the estimate of the percentage of persons missed by the PES but found by the Census, and the percentage of persons missed by the Census but found by the PES, can be used to construct estimates of the percentage of persons missed by both the PES and the Census.

Despite efforts to maintain independence, the likelihood of a person being missed in the PES may be related to whether they were missed in the Census. This would result in a 'correlation bias' in the PES estimates. To minimise this bias, sample weights can be adjusted during the PES estimation process to take account of the fact that different groups have a different likelihood of being missed.

Changes to the 2016 Census and the 2016 PES

CHANGES TO THE 2016 CENSUS AND THE 2016 PES

There are a number of key changes to collection operations for the 2016 Census that may result in a change to the overcount and undercount, as measured by the PES. These include:

- a new digital-first approach to collection operations, promoting higher adoption of an online Census; and
- use of an Address Register as the basis for mail-out of Census materials.

The 2016 PES has struck a balance between making essential adaptations for the new 2016 Census model, while preserving comparability with the 2011 PES where possible, to ensure consistency of population estimates over time.

New Approach to the 2016 Census

NEW APPROACH TO THE 2016 CENSUS

COLLECTION OPERATIONS

The ABS has developed a new digital-first approach to the 2016 Census that will provide a faster, more efficient, more environmentally-friendly Census that is easier for people to complete. The new approach will overcome difficulties in recruiting field staff and at the same time take advantage of new technologies.

The new approach changes the way Census materials are delivered to householders and information is returned to the ABS. These changes were designed taking into account international best practices in Census procedures, and building on the Australian public's increasing access to and use of the internet, and their willing support of the Census.

Under the traditional Census method used for the past 100 years, forms were delivered by hand to every dwelling. The new delivery approach removes the need for Census Field Officers to visit every dwelling. Instead, approximately 80% of dwellings across Australia will, in the first instance, be mailed information which includes a unique login number for the online form. The online form is designed to be used easily and securely on a variety of devices from smart phones to desktop computers. Those residents who do not wish to complete their form online will be able to request a paper form, which they can complete and mail back in a provided prepaid envelope.

For dwellings that have not responded by a specified date, reminder letters will follow the initial correspondence. Census Field Officers will then only visit dwellings that have still not responded.

In the remaining areas of Australia, a more traditional delivery approach will be used. In these areas, Census Field Officers will deliver materials to each dwelling, enabling residents to either complete their form online or mail back a paper form. In these areas, the Field Officers will attempt to make contact with residents when dropping off the form. Census Field Officers will then only make further visits to dwellings that have not responded.

It is expected that about two-thirds of Australians will respond online to the 2016 Census, doubling the online response rate in 2011 of 33%.

ADDRESS REGISTER

Central to the new Census delivery procedures is the ABS Address Register. The ABS has developed this register as the central source of addresses used in the collection of information. The main input to the register is the Geocoded National Address File (G-NAF), with continuing supplementation from other available address sources and from field work undertaken by ABS officers. Each record listed on G-NAF has:

- an address;
- geocode information (latitude and longitude coordinates); and
- land use details.

Residential addresses from the register will be used for the mail-out of unique Census login numbers and other correspondence. They will also be used by Census Field Officers to follow up dwellings that have not returned a form. In areas where the register is considered to be of a lesser quality, necessitating delivery of Census materials by Census Field Officers, the officers will record the addresses of the dwellings they visit, for inclusion in the register.

Adapting the 2016 PES

ADAPTING THE 2016 PES

PES ENUMERATION

2016 PES has two key changes to adapt to the new Census model:

- the 2016 PES sample size has been increased by 20%, which is sufficient to cover the expected increase in sampling error due to population growth and new overcount and undercount effects; and
- the 2016 Census follow-up phase starts and finishes several weeks later than the 2011 Census follow-up phase, and so the 2016 PES also starts and finishes several weeks later, immediately following the Census follow-up phase.

The 2016 PES has prepared a telephone interview (TI) option in addition to the traditional face to face interview. The TI option can be initiated by the respondent using contact information provided in the Primary Approach Letter. The availability of another mode of collection for PES is expected to improve response rates, especially from those who would be more likely to participate by telephone than face-to-face.

CONTACT SECTOR AND NON-SECTOR

For the purposes of the PES estimation, Census dwellings deemed occupied on Census night are divided into a Contact Sector (dwellings for which a Census form was received before the commencement of the PES enumeration) and a Non-Contact Sector (where no Census form was received, or a Census form was received after the commencement of the PES enumeration and is classified as a 'late return').

For the Contact Sector, it is usual for the PES to measure a net undercount of persons, with the number of persons missing from Census forms generally exceeding the number of

persons counted twice on two Census forms. In 2011, the size of the Contact Sector was around 96.5% of dwellings, and the PES measured a net undercount of 635,000 persons missed from these dwellings.

For the Non-Contact Sector, it is usual for the PES to measure an overcount of persons, mainly due to imputation by Census for non-responding dwellings thought to be occupied. In 2011, the size of the Non-Contact Sector was around 3.5% of dwellings, and the PES measured an overcount of 260,000 persons in these dwellings.

International experience suggests that there will be new overcount and undercount mechanisms, because the Address Register is expected to have both over and undercoverage. A possible scenario under the new Census model is for the PES to report a net overcount if Address Register overcoverage leads to increased imputation in the Non-Contact Sector. Improvements in collection and linking of dwelling and address information have been made for the PES to ensure that new overcount and undercount mechanisms can be correctly measured and adjusted to give accurate population estimates.

LINKING AND ESTIMATION

There are no fundamental changes to how the PES selects the sample of dwellings, links records to Census and estimates the Census night population. While it is anticipated that the overcount and undercount may change with the introduction of the new Census model in 2016, it is expected that the PES will give an accurate estimate of the Census night population, as it did in 2011. In particular:

- area-based PES sample selection will be used to complement and contrast the list-based Census model, and ensure statistical independence through sourcing dwelling lists with independent procedures;
- Automated Data Linking (ADL) of persons will be comparable to the 2011 Census and the 2011 PES;
- Prediction Regression Estimation (PREG) will be comparable to the 2011 Census and the 2011 PES. Reviews undertaken by the ABS have confirmed that the PREG methodology can adjust for both overcount and undercount when estimating the adjusted Census night population of Australia; and
- the identification question for Aboriginal and Torres Strait Islander peoples in the 2016 PES will be the same as that used in the 2011 PES.

Methodology of the 2016 PES

METHODOLOGY OF THE 2016 PES

The following sections describe the methodology of each component of the PES statistical process, including:

- scope;
- sample design;
- collection methods;
- questionnaire design;
- address coding;
- automated data linking;
- the PES match and search system; and
- estimation.

Scope

SCOPE

The scope of the PES is similar to the scope of the Census, which is every person present in Australia on Census night (with the exception of foreign diplomats and their families). Ideally the PES would sample from all people who were or should have been in the scope of the Census, but for practical reasons there are a number of areas, dwellings and people excluded or not able to be covered by the PES. Of the people present in Australia at the time of the PES, the following are not included in the PES scope:

- foreign diplomats and their families;
- people in non-private dwellings such as hotels, motels, hospitals and other institutions;
- homeless people (as the sample selected in the PES is based on the selection of dwellings);
- overseas visitors who were not in Australia on 9 August 2016 (Census night);
- babies born after 9 August 2016; and
- people in Cocos (Keeling) Islands, Christmas Island, Australian Antarctic Territory and Jervis Bay Territory.

The PES does not obtain information about people who died between Census and the PES, however it does obtain information about Australian residents who are overseas during the PES enumeration period and who departed after the Census, provided that they usually live with people remaining in Australia.

REMOTE AREAS AND DISCRETE COMMUNITIES

Remote areas and Discrete Communities were first included in the PES in 2006. Prior to this, they were excluded from the PES because of operational issues associated with enumerating them. In addition, it was considered difficult in the past to implement procedures to ensure that independence of the PES from the Census was not compromised by using the same local contacts for both Census and PES.

In 2006 and 2011, the risk to operational independence was effectively managed through interviewer training and the procedures put in place for field staff. Enumeration of Discrete Communities will occur in 2016 across August, September and October.

NON-PRIVATE DWELLINGS

Non-private dwellings are establishments which provide predominantly short-term accommodation for communal or group living, and often provide common eating facilities. They include hotels, motels, hostels, hospitals, religious institutions providing accommodation, educational institutions providing accommodation, prisons, boarding houses and short-stay caravan parks. Non-private dwellings each comprise a number of dwelling units. About 2% of the Australian population live in non-private dwellings. Non-private dwellings have previously been excluded from the PES because:

- a high number of these dwelling units are found to be vacant;
- many of the people who are contacted in these dwelling units are not in scope of the PES, for example overseas visitors who were not in Australia on 9 August;
- many of the people usually live in a private dwelling and could be selected at their usual residence;
- the quality of information is often poor, especially in institutions, where information cannot be provided by individuals themselves and so is collected from administrative lists or from staff;
- enumeration suffers from a lack of independence from the Census, because the administrative lists used by PES interviewers to collect data about institutionalised people are the same as those used by Census staff; and
- the cost of enumerating them is relatively high compared to private dwellings, because of the additional time and effort required.

An investigation was conducted into the possible inclusion of non-private dwellings in the 2006 PES. The investigation found that non-private dwellings were likely to have high levels of sample loss and non-response, which meant their inclusion in the PES would not be cost-effective. Non-private dwellings were therefore excluded from the 2006 PES and 2011 PES.

For 2016 PES, the inclusion of non-private dwellings was revisited, specifically the subset of long-stay non-private dwellings such as educational institutions with accommodation, because interviewers have reasonable prospects of collecting good quality data directly from usual residents. However, this subset represent only about one-third of the non-private dwelling population and, coupled with the changes to Census enumeration, the decision was taken to minimise the number of non-essential changes to the PES. Non-private dwellings are therefore excluded from the 2016 PES.

COVERAGE RULES

The PES coverage rules are designed to give each in-scope person in the population a single chance of selection in the survey by associating each person with one and only one dwelling. The coverage rules are implemented by asking a series of questions in the PES interview. The questions cover topics such as where each person usually lives and whether they will be staying at the PES dwelling on a particular reference night.

Structuring the questionnaire so that each person has a single chance of selection in the PES ensures that data from the sample can produce estimates that are representative of the whole population.

Sample Design

SAMPLE DESIGN

In the PES, private dwellings (houses, flats, etc.) and Discrete Communities are separately identified and sampled. In total, about 43,000 dwellings are expected to respond to the PES. This represents an increase of 20% on the size of the 2011 PES sample. The sample includes approximately 600 dwellings selected from 33 Discrete Communities in NSW, Qld, SA, WA and NT.

Table 1 shows the expected number of fully responding households for each state and territory. It is important to note that, due to sample loss and other non-response, the number of dwellings selected in the sample will be greater than shown below. Based on the results of other ABS surveys, the rate of sample loss is expected to vary between different components of the sample, which is taken into account.

Table 1: Expected number of fully responding households

	Total no.
New South Wales	9 400
Victoria	8 200
Queensland	7 900
South Australia	4 700
Western Australia	5 200
Tasmania	2 500
Northern Territory	3 400
Australian Capital Territory	1 700
Australia	**43 000**

This sample size is expected to be sufficient to provide estimates of the net overcount or net undercount rate with the following standard errors:

- persons at Australia level - standard error to be less than 0.2 percentage points;
- persons at state/territory level - standard error to be less than 0.6 percentage points, for state/territories larger than 500,000 people;
- persons at state/territory level - standard error to be less than 3,000 persons for states/territories smaller than 500,000 people*;
- Indigenous persons at Australia level - standard error to be less than 2.0 percentage points;
- males and females at Australia level - standard error less than 0.3 percentage points;
- persons at Australia level by 5 year age brackets - standard error less than 0.6 percentage points.

* NT and ACT had populations below 500,000 as at December 2014.

GENERAL POPULATION SAMPLE

For all general population areas other than Discrete Communities, private dwellings are selected through a stratified, multistage cluster design from the private dwelling framework of the ABS Population Survey Master Sample. The geographic area used for selection is comprised of mesh blocks from the Australian Statistical Geographical Standard (ASGS). Areas are grouped into strata based on locality, population density, remoteness, projected population growth and socio-economic characteristics. Each private dwelling within a stratum has the same probability of selection.

The 2011 PES used larger sampling fraction for areas with larger proportions of households

with Aboriginal and Torres Strait Islander persons in residence, based on data from the 2006 Census. The aim of this was to improve estimates of Indigenous overcount and undercount. However, a review of the 2011 PES sample design showed that most of the improvement of Aboriginal and Torres Strait Islander estimates was achieved by increases to the overall sample size.

DISCRETE COMMUNITIES SAMPLE

For the purposes of sampling, Discrete Communities are grouped into 'sets' comprising main communities and their associated outstations. The selection of main communities is undertaken with probability proportional to the size of the set. The aim is to select a representative sample while also considering cost constraints, reasonable interviewer workloads and expected sample size.

For communities selected in the sample, selection of dwellings within the community is performed using systematic equal probability sampling.

A selection of outstations associated with each selected main community is also included in the sample. Each outstation has an equal chance of selection and, once selected, all dwellings within the outstation are enumerated. Occasionally, an outstation is not enumerated if it is difficult to access from the main community, resulting in a small loss of coverage.

The 2016 PES has selected a sample of Discrete Communities similar in size to that of 2011 PES in Qld, SA, WA and NT, with the addition of a small sample in NSW. For each state, the sampling fraction for Discrete Communities is 2 to 3 times the sampling fraction for the rest of the state.

Collection Methods

COLLECTION METHODS

FIELD PROCEDURES

Various strategies have been devised for the enumeration of the PES. Where possible, standard ABS survey procedures are used when enumerating private dwellings and Discrete Communities. However, these procedures are modified where necessary in Discrete Communities to take account of language and cultural issues. While question wording and collection methodology may be modified in some cases, attempts will be made to ensure underlying concepts remain the same across both components of the sample.

GENERAL POPULATION SAMPLE

Specially trained PES interviewers will collect data through interviews, starting around seven weeks after Census night. All private dwellings will be enumerated using Computer Assisted Interviewing (CAI). Interviews will be conducted with any responsible adult of the household who will be asked to respond on behalf of all household members. Most interviews will be conducted face-to-face, however respondents will have the option to register instead for a telephone interview, using instructions on a Primary Approach Letter (PAL) delivered by mail to the selected dwellings approximately one week before the start of interviewing. The introduction of telephone interviewing will improve the effectiveness of enumerating the large 2016 PES sample in a short time frame, as well as giving respondents greater choice

and convenience.

The wording of the identification question for Aboriginal and Torres Strait Islander people in the PES survey will be the same as that used in 2011.

In each Census there are always dwellings for which Census forms have not been returned within the required timeframe. For this reason, intensive Census follow-up procedures are employed at the end of Census collection. To allow sufficient time for Census follow-up and to minimise potential overlap, 2016 PES enumeration will begin in late September, just after the end of the Census follow up phase, and will continue until late October.

DISCRETE COMMUNITIES SAMPLE

In Discrete Communities, the primary collection method will be a customised questionnaire using Computer Assisted Interviewing (CAI). The PES will be conducted by specially trained ABS staff with the assistance of facilitators recruited from within the community. The facilitator's role is to assist in establishing rapport with respondents, to assist the interviewer in identifying residents of the selected households, and to interpret where necessary.

To strengthen the independence of the Census and the PES, efforts will be made to recruit facilitators who were not involved as Census interviewers. Where this is not possible, Census interviewers will only act as PES facilitators at dwellings where they did not interview during the Census. If there is no option but to use a facilitator who conducted a Census interview at the same dwelling, the PES interviewer will ask the facilitator to introduce the interviewer to the household, but otherwise to take no part in the interview. Alternatively, where acceptable to the community, the PES interviewer may enumerate dwellings without the assistance of a facilitator.

Census enumeration in Discrete Communities will take place between late July and early September, and PES enumeration will take place as soon as practicable after. To avoid any overlap, PES enumeration will only begin in these communities once it has been established that all Census field activities have been completed. As a result, PES enumeration of the Discrete Communities sample will be staggered, to align with the completion of Census field operations in each of the communities. PES enumeration in some communities may start as early as August, and all communities in the sample will be completed by late October.

INDEPENDENCE BETWEEN CENSUS AND PES

A risk to the independence of the Census and the PES as some people may be prompted to return their Census forms following receipt of the PES primary approach letter or the arrival of the PES interviewer. Any Census form received after the receipt of the Primary Approach Letter will be flagged as a 'late return'. The treatment of late returns is explained in the Estimation section of this paper.

The independence of the Census and the PES will also be strengthened by strategies designed to minimise the overlap of Census collectors and PES interviewers. These include:

- interviewers wanting to apply for District Manager or Area Supervisor positions in the Census must preclude themselves from the interviewer panel for the PES;
- PES interviewers may also work as Census collectors but they must enumerate a different area;
- any person who has worked in an area of Census processing related to dwelling or population counts may not also work on PES processing; and
- any person who has been involved in Census field collection in any capacity may not also work on PES processing.

Questionnaire Design

QUESTIONNAIRE DESIGN

GENERAL POPULATION QUESTIONNAIRE

The PES questionnaire collects personal details (name, sex, date of birth, age, relationship in household, marital status, country of birth and Indigenous status) to facilitate matching of PES person records to Census person records, and to allow accurate overcount and undercount estimates to be generated for age and sex categories, and Indigenous status. For each person in the household, the PES also asks:

- whether they were included on a Census form (and if so, where);
- whether they could have been included on a Census form at other addresses (and if so, where); and
- where they stayed on Census night.

The different addresses collected in the PES are used to search Census records to determine the number of times each PES respondent was included in the Census. Visitors to households included in the PES are also asked for their address of usual residence. In addition to questions relating to Census night, the PES collects a small amount of information on dwelling tenure and structure.

DISCRETE COMMUNITY QUESTIONNAIRE

As in 2011, a specially designed questionnaire will be used in Discrete Communities. A number of questions in the private dwelling PES questionnaire are not considered applicable to people living in Discrete Communities and have been adapted to ensure that information is collected in the most culturally appropriate manner. For example, it is more common for Aboriginal and Torres Strait Islander persons in the community to have more than one name by which they may be known. This tailored questionnaire ensures that sufficient information is collected to effectively match persons, taking into account the potentially different nature of matching.

A community-level questionnaire will also be asked of the community contact or council officer of Discrete Communities. Information collected, such as whether any significant event (for example a sports carnival) may have occurred at the time of the Census, is expected to help with respondent recall and assist in the completion of the individual questionnaires.

DATA CAPTURE

Collected data will be converted into files containing dwelling-level information (address, number of people, response status, etc.), person-level information (name, age, sex, marital status, etc.), search addresses (where the person may have been included on a Census form), and interviewer comments (e.g. to assist in dwelling and person matching). The files then undergo basic data cleansing, amendments, derivations and quality assurance procedures.

Address Coding

ADDRESS CODING

Address information is essential for matching people between PES and Census data. Matching a person between the data sources is facilitated by the location of the dwelling where they were enumerated in PES, where they were included on a Census form and where they may have been included on a Census form. PES has two purposes that require the processing of address information:

- "Geocoding" is the process whereby an address is located and assigned to geographic information such as a Coordinate (e.g. Latitude & Longitude), a Mesh Block (MB) or a Statistical Area Level 1 (SA1) boundary. Geocoding is required for automated datalinking of persons. PES requires the MB and SA1 location to search for person matches during automated data linking for all PES enumeration addresses (where the PES interview took place) and all 'search addresses' (other locations that respondents may have been present on Census night, or recorded on a form on Census night). Geocoding is relatively robust to errors (eg character substitution or form scanning errors) in the address text as it needs only to identify the locality and not a specific address or dwelling; and
- "Address text parsing & matching" is the process whereby an address is matched to a specific entry on an existing address index such as the Address Register. PES benefits from identifying an exact dwelling match where possible, in order to identify a specific Census form to examine during clerical review. Address text matching is more precise than geocoding, seeking an exact address match rather than the locality, but is more susceptible to errors or missing entries in the address list.

Automated Data Linking

AUTOMATED DATA LINKING

Automated Data Linking (ADL) refers to the use of automated data linking processes to determine possible links between Census and PES data before any clerical matching process begins. ADL was introduced into the PES in 2011 for the first time and will be used again for 2016.

The approach to data linking is to generate large numbers of candidate links and then use a process of elimination to filter down to genuine matches. ADL employs probabilistic data linking techniques (Fellegi et al 1969), using a range of personal and address characteristics, to evaluate the chance that a PES record and a Census record are the same person. ADL therefore provides the opportunity to match persons that would have been too difficult to match using only clerical match and search procedures. The key gains in matching effectiveness and efficiency provided by ADL include:

- the ability to conduct a more comprehensive search for PES respondents than was possible from previous clerical matching processes;
- the ability to locate PES respondents at Census night addresses, regardless of whether PES respondents recall all the locations they may have been recorded on a Census form; and
- a reduced requirement for clerical matching resources.

ADL uses Freely Extensible Biomedical Record Linking (FEBRL) software developed at the Australian National University. A number of different linking runs will be used to compare PES and Census records, each focused on a slightly different combination of name, addresses and demographic variables.

Potential links are assessed by assigning weights that reflect the level of agreement on selected data items from the two records. Large positive weights indicate probable matches, while large negative weights are observed for probable non-matches.

Important to the effective use of ADL is a series of processes that are run after ADL output is obtained. The Collect, Analyse, Reduce, De-duplicate and Systematise (CARDS) process identifies and rates the most plausible links from each ADL run for all PES respondents. The process then combines the links from all ADL runs and removes any duplicates.

While ADL is the next step in the evolution and continual improvement of PES processing, it is important to note that ADL cannot entirely replace the clerical decision-making process that has previously been at the core of PES processing. Clerical judgment will always be required to resolve the more complex or ambiguous cases and be used as a means of quality assuring automated processes.

To clerically confirm or reject potential person links identified by ADL, it is efficient to handle links for all persons in the same PES dwelling, so as to examine the entire household and their corresponding Census form at the same time. For this reason the CARDS process concludes by creating dwelling links and assigning Dwelling Link Ratings (DLR), to determine which PES dwellings will be examined by the PES Match and Search System (MSS), and which can be confirmed as links without further review. Person-level links are grouped together according to the dwelling they were found in PES, and dwelling-level links are derived and rated on the basis of the number and quality of person-level links identified between the PES and Census dwellings.

PES uses three Dwelling Link Ratings:

- **Platinum** links are of sufficiently high quality that they are confirmed immediately and do not require clerical review;
- **Silver** links are of moderate level quality, for example high quality links between some persons in a dwelling and not others, and require clerical review to confirm or reject the quality of the link;
- **Tin** links are of low level quality and unlikely to be true matches, so clerical review is required to search for matches for these dwellings and persons without the assistance of ADL.

All PES dwellings with either Silver or Tin links are sent for clerical review. A small percentage of Platinum links are also clerically reviewed for quality assurance purposes.

The PES Match and Search System

THE PES MATCH AND SEARCH SYSTEM

The Match and Search System (MSS) is the main PES clerical review facility. The MSS allows processing staff to search, view, compare, record matches, and reject non-matches between PES and Census data. PES processing staff use the MSS to record matches of

dwellings and people between PES and Census, and to search for people on Census forms at alternative 'search addresses' specified by a PES respondent.

There are two modes of use for the MSS: evaluation of a candidate link provided by the ADL process (either Platinum or Silver), or clerical search for a link in the absence of a good candidate from ADL (Tin). These are done in two phases, so that any case where an ADL candidate link is rejected can then revert to the full clerical search.

The first phase of MSS processing involves confirming whether the ADL-identified dwelling link is correct. Once a dwelling link is confirmed, the Census person records for that dwelling are compared with the PES person records. The information compared includes name, sex, date of birth, age, marital status, Indigenous status and country of birth. The extent to which each of these items of information is the same in both the PES and the Census determines the ADL match status of the pair and the level of match.

Once all ADL links have been reviewed, the second phase in MSS processing is to conduct an intensive clerical search for persons not matched as a result of ADL information. This is done by searching in geographical areas for addresses provided by respondents during the PES interview, in order to locate possible Census forms.

MSS QUALITY ASSURANCE

To ensure the accuracy of MSS processing, quality assurance (QA) procedures are used in the match and search process whereby all PES records that were clerically reviewed once are processed a second time by a different operator. Where the initial and QA processing outcomes correspond, the initial match status is accepted.

Where there is a discrepancy between the initial match status and the QA match status (either on dwelling match or person match code), the records are flagged for adjudication by a senior officer who reviews all information and determines which is correct. If both initial and QA records are deemed to be inaccurate, the adjudicator reprocesses the record.

MSS PROCESSING FOR DISCRETE COMMUNITIES

MSS processing for Discrete Communities will be similar to that for the private dwelling component, with an additional approach that accounts for the fact that some Indigenous Australians are associated with more than one dwelling and move between these dwellings on a regular, seasonal or random basis (Kinfu, 2005). This means that a number of people may not be staying in the same dwelling during the PES that they were during the Census.

For this reason, the matching process in Discrete Communities involves searching the whole community for a person match, rather than just searching within a single dwelling. Person matching in Discrete Communities will use the same rules for determining a match as in the private dwelling component, but will use up to two alternative names by which the person is known when matching on name.

Estimation

ESTIMATION

The PES interview determines whether each person in the sample should have been counted in the Census, and the linking and matching process determines how many times

each person was actually counted in the Census. The PES estimation process combines and weights these data to produce an estimate of the number of people who should have been counted in the Census. This difference between this and the number of people who were actually counted in Census is the estimate of net overcount or net undercount.

WEIGHTING FOR THE PES

Weighting is the process of adjusting results from a sample survey to infer results for the total in-scope population. The weight can be considered an indication of how many people or dwellings are represented by those in the sample.

To derive estimates for the entire population in the scope of the PES, expansion factors (weights) are applied to the sample responses. The weighting method ensures that estimates conform to known population totals by age, sex and geographic area. It considers population information such as the number of dwellings of different characteristics that were located in Census and the number of people with different characteristics that were counted in the Census. This reduces sampling variability and compensates for any under-enumeration or non-response in the survey.

The underlying principle adhered to in the weighting is that any person in the sample will receive the same weight, regardless of whether or not they were counted in the Census. In a simplistic example the weights of PES people that were counted in the Census should be adjusted so that the sum of the weights is equal to the number of people counted in the Census. The weights applied to PES sample respondents who were *not* counted in Census should be adjusted in the same way as the weights of those that *were* counted. The sum of the weights of all PES respondents will then be an estimate of the number of people that should have been counted in the Census.

The weighting classes classify the population by:

- state or territory and geographic region;
- sex;
- age;
- country of birth;
- marital status;
- Aboriginal and/or Torres Strait Islander origin and
- whether selected from a Discrete Community.

The weight adjustment applied to a person depends only on the information reported in the PES.

As a final step in weighting, weights are adjusted so that the PES estimates represent people in non-private dwellings as well as private dwellings. This final step uses weighting class categories of state or territory, region, age and sex only, as information on other items is not considered reliable for non-private dwellings.

The method used for adjusting weights to meet population totals is the Prediction Regression (PREG) estimator. PREG estimation is able to take into account people counted multiple times in the Census and people who give different responses between PES and Census. A detailed description of the PREG estimator can be found in Research Paper: An Estimating Equation Approach to Census Coverage Adjustment, May 2007 (cat. no. 1351.0.55.019).

How Overcount and Undercount Contribute to the Estimated Resident Population

HOW OVERCOUNT AND UNDERCOUNT CONTRIBUTE TO THE ESTIMATED RESIDENT POPULATION

BACKGROUND

The estimated resident population (ERP) is the official estimate of the population calculated quarterly by the ABS. Accurate estimates of the Australian resident population are essential for the allocation of seats in the Federal House of Representatives as well as distributing Commonwealth funding to states and territories.

These population estimates are also used for demographic, social and economic studies. The validity of these population estimates depends on a number of factors, one of which is the accurate measurement of net overcount or net undercount in the Census.

Following each Census, the ERP for 30 June of the Census year is rebased using the Census data, adjusted by the PES.

Calculation of ERP

CALCULATION OF ERP

Initially, a count of Australian residents who were in Australia on Census night is obtained from the Census. This Census count is adjusted on the basis of net overcount or net undercount, to account for people being counted more than once and people being missed in the Census. The basis of the estimate of net overcount or net undercount is the PES.

ERP calculations then make an adjustment for Australian residents who were temporarily overseas on Census night. An estimate of this number is made using data from completed passenger cards, visa and passport information obtained from the Department of Immigration and Citizenship and these people are added into the Australian resident population.

The final step in calculating ERP is to backdate to 30 June of the Census year. This is achieved by adding the deaths and subtracting the births and net overseas migration which occurred between 1 July and the Census date. Table 2 shows the components used to calculate ERP for Australia from the 2011 Census.

Table 2. Components of ERP, Australia, 2011

Components of ERP (a)	Persons '000
Census count, actual location	21 727.2
less Overseas visitors	-219.4
Census count, place of usual residence	21 507.7
plus Net undercount (b)	386.4
plus Residents temporarily overseas	484.7
ERP (a) as at 9 August 2011	22 378.8
less Births (1 July to 9 August 2011)	-32.5
plus Deaths (1 July to 9 August 2011)	17.7

(a) Preliminary ERP.
(b) Includes demographic adjustments.

Information on the calculation of the ERP for 30 June 2016 based on the 2016 Census will be reported in Australian Demographic Statistics, December quarter 2016 (cat. no. 3101.0), due for release on 21 June 2017. Additional information can also be found in Population Estimates: Concepts, Sources and Methods, 2009 (cat. no. 3228.0.55.001).

DEMOGRAPHIC ADJUSTMENTS

While the PES identifies people and dwellings missed in the Census, the extent to which some people are missed in both the Census and the PES may not be fully reflected in PES estimation, which would result in a correlation bias. As in any survey, the PES is also subject to sampling and non-sampling error.

To offset the impact of correlation bias and survey error, population estimates derived from the PES are further refined using demographic adjustments based on three sources of independent population information: the National Demographic Data Bank, Medicare enrolment numbers, and the estimated resident population based on the previous Census. These sources have different strengths and weaknesses, but where the data are considered to be most reliable, they are used for comparison with PES adjusted age and sex population distributions, and for possible minor adjustments to population estimates.

The National Demographic Data Bank is a population database maintained by the ABS using administrative data (notably births, deaths, and overseas arrivals and departures). The database is independent of Census data and contains population data back to 1925. For the 2006 PES, these data were considered to measure age-sex totals well up to about age 35, after which there were some concerns about pre-1970 international migration data. Sex ratios derived from these data are considered most reliable for ages under 28 years.

Enrolment data from Medicare (the Australian government health rebate system) are considered a good source for calculating sex ratios, but less reliable for age-sex population totals. Age-sex totals are least reliable among the older ages where people may remain enrolled in Medicare after their death until Medicare Australia is notified and the record updated.

About this Release

Explains the methodology to be used in conducting the 2016 Census Post Enumeration Survey, including changes since the 2011 survey.

Explanatory Notes

Bibliography

BIBLIOGRAPHY

ABS 2007, Research Paper: An Estimating Equation Approach to Census Coverage Adjustments, May 2007, cat. no. 1351.0.55.019, ABS, Canberra.

ABS 2009, Population Estimates: Concepts, Sources and Methods, 2009, cat. no. 3228.0.55.001, ABS, Canberra.

ABS 2011, Information Paper: Measuring Net Undercount in the 2011 Population Census, 2011, cat. no. 2940.0.55.001, ABS, Canberra.

ABS 2012, Census of Population and Housing - Details of Undercount, 2011, cat. no. 2940.0, ABS, Canberra.

ABS 2013, Australian Demographic Statistics, December quarter 2012, cat. no. 3101.0, ABS, Canberra.

ABS 2015, Census of Population and Housing: Nature and Content, Australia, 2016, cat. no. 2008.0, ABS, Canberra.

Kinfu, Y., 2005, 'Spatial mobility among Indigenous Australians: Patterns and determinants', Demography and Sociology Program - Working papers in Demography No. 97, Australian National University, Canberra.

Fellegi, I. P. and Sunter, A. B., 1969, 'A Theory for Record Linkage', Journal of the American Statistical Association, Vol 64, No 328, American Statistical Association.

Glossary

GLOSSARY

Address Text Matching
The process whereby an address is matched to a specific entry on an existing address index such as the ABS Address Register.

Automated Data Linking (ADL)
Automated linking processes used to determine possible links between Census and PES data, before any clerical matching process has begun. It employs a probabilistic linking method that uses a range of personal and address characteristics to evaluate the chance that a PES and Census record are the same individual.

Collect, Analyse, Reduce, De-duplicate and Systematise (CARDS)
A series of processes which prioritises and organises ADL output for use by the MSS system.

Computer Assisted Interviewing (CAI)
A method of data collection whereby responses are recorded directly into an electronic questionnaire on a notebook computer.

Correlation bias
A bias arising when people who were not counted in the Census are more likely to be missed in the PES than people with similar characteristics such as age, sex and Indigenous status who were counted in the Census.

Coverage

The coverage is the actual population of persons within the scope of a survey about which data can actually be collected. As it is not always possible to collect data from all persons, some persons may be in scope but not in coverage. For the quality of the survey estimates, it is desirable that the survey coverage matches as closely as possible the survey scope. Coverage rules are generally applied in all surveys to ensure that each person is associated with only one dwelling, and hence has only one chance of selection.

Contact Sector

The set of all dwellings for which a Census form was received before the commencement of the PES enumeration. PES uses Census records from these dwellings during weight adjustment to improve the accuracy of estimates.

Discrete Community

A geographic location, bounded by physical or legal boundaries, and inhabited or intended to be inhabited predominantly by Aboriginal and Torres Strait Islander people, with housing or infrastructure that is either owned or managed on a community basis.

Dwelling Link Rating (DLR)

A system of classifying the quality of links for a dwelling of PES respondents to determine what level of clerical review is required. The three link ratings are Platinum, Silver and Tin.

Estimated Resident Population (ERP)

The official measure of the population of Australia based on the concept of residence. It refers to all people, regardless of nationality or citizenship, who usually live in Australia, with the exception of foreign diplomatic personnel and their families. It includes usual residents who are overseas for less than 12 months of the preceding 16 months. It excludes overseas visitors who are in Australia for less than 12 months of the preceding 16 months.

ERP rebasing

Following each Census of Population and Housing, ABS rebases its official estimates of population - the Estimated Resident Population (ERP) series. This provides the basis for population estimates until the next Census. The rebased Census-year ERP together with other Census information and data on intercensal components of population change allows ERP revisions to be made back to, but not including, the previous Census.

Freely Extensible Biomedical Record Linking (FEBRL)

Probabilistic linking software developed at the Australian National University used for PES automated data linking activities.

Geocoding

The process whereby an address is assigned a geographic classification.

General Population sample

All PES private dwelling sample selections other than those from discrete communities.

Household

A household comprises either a single person, a family unit or a group of people within a dwelling who know each other and are able to provide information about each other. Usually there is only one household in each private dwelling, but separate households may be created if certain conditions apply.

Imputation

A statistical process for predicting values where no response was provided to a question and a response could not be derived.

Imputed dwelling (in Census)
A dwelling which is determined to be occupied in the Census and where Census data is imputed because no Census form was received.

Intercensal Difference
The net revision (upward or downward) of the Estimated Resident Population as a result of rebasing.

Late return
A Census form which was returned after the start of PES enumeration.

Match and Search System (MSS)
The main PES clerical review facility, which allows processors to search, view, compare, and record matches between PES and Census data.

Non-contact sector
The set of all dwellings where no Census form was received, or a Census form was received after the start of the PES enumeration and is classified as a 'late return'. To preserve statistical independence, PES cannot use information from either 'late returns' or imputed dwellings from the Census when adjusting weights, but they do contribute to the undercount and overcount equations during tabulation.

Non-private dwelling
An establishment which provides a communal type of accommodation, such as a hotel, motel, hospital or other institution.

Non-sampling error
Error arising from inaccuracies in collecting, recording and processing the data. Every effort is made to minimise non-sampling error by the careful design of questionnaires, intensive training and supervision of interviewers, and efficient data processing procedures. Non-sampling error also arises because information cannot be obtained from all persons selected in the survey and data are imputed based on assumption about the non-respondents.

Outstation
A discrete community that has a population of fewer than 50 people AND is administered by, or linked to, an organisation such as a Resource Agency or larger parent discrete community for the provision and maintenance of services.

Private dwelling
A residential structure which is self-contained, owned or rented by the occupants, and intended solely for residential use. A private dwelling may be a flat, part of a house, or even a room, but can also be a house attached to, or rooms above, shops or offices.

Sampling error
Sampling error occurs because data were obtained from a sample rather than the entire population. One measure of the likely difference resulting from not including all dwellings in the survey is given by the standard error. There are about two chances in three that a sample estimate will differ by less than one standard error from the figure that would have been obtained if all dwellings had been included in the survey, and about nineteen chances in twenty that the difference will be less than two standard errors.

Scope
The scope of a survey is the population about which information is to be collected.

Search address
An address where a person was reported to be staying on Census night or where a person may have been included on a Census form. PES processing attempts to locate all Census forms for each search address, in order to determine the number of times (if any) a person enumerated in the PES was included on a Census form.

Standard error
A measure of the sampling error.

Statistical Independence
A mathematical assumption that underlies PES estimation. Statistical independence requires both population independence (that people who were not counted in the Census are no more likely to be missed by the PES than people who were counted) and operational independence (that that Census operations do not influence PES operations and vice versa).

Systematic Equal Probability Sampling
A method of sampling from an ordered list by choosing one element, skipping past several others, choosing the next element and so on. The skip determines how many to take, for example a skip of 10 means take 1 in 10. For selecting dwellings from a block, this means compiling a list of dwellings, choosing at random the first dwelling to select, then applying the skip repeatedly to the end of the list.

Unoccupied dwelling (in Census)
A structure built specifically for living purposes which is habitable but the Census Collector was certain was unoccupied on Census night.